REVISITING THE SANDERS-FREIMAN-RUZSA THEOREM IN $\mathbb{F}_p^n$ AND ITS APPLICATION TO NON-MALLEABLE CODES

DIVESH AGGARWAL AND JOP BRIËT

ABSTRACT. Non-malleable codes (NMCs) protect sensitive data against degrees of corruption that prohibit error detection, ensuring instead that a corrupted codeword decodes correctly or to something that bears little relation to the original message. The *split-state model*, in which codewords consist of two blocks, considers adversaries who tamper with either block arbitrarily but independently of the other. The simplest construction in this model, due to Aggarwal, Dodis, and Lovett (STOC'14), was shown to give NMCs sending k-bit messages to $O(k^7)$-bit codewords. It is conjectured, however, that the construction allows linear-length codewords.

Towards resolving this conjecture, we show that the construction allows for code-length $O(k^5)$. This is achieved by analysing a special case of Sanders's *Bogolyubov-Ruzsa theorem* for general Abelian groups. Closely following the excellent exposition of this result for the group $\mathbb{F}_2^n$ by Lovett, we expose its dependence on p for the group $\mathbb{F}_p^n$, where p is a prime.

1. INTRODUCTION

1.1. **Non-malleable codes.** *Non-malleable codes* (NMCs) aim to protect data when it is subjected to the kind of corruption that renders reliable error correction and detection impossible. The defining feature of such codes is that an adversary who tampers with a codeword will have little control over what it decodes to. Despite having appeared only recently [DPW10], these codes already emerged as a fundamental object at the intersection of coding theory and cryptography, such as in the construction of non-malleable commitment schemes [CGM+15, GPR15] and non-malleable encryption schemes [CMTV15, CDTV16]. The study of non-malleable codes falls into a much larger cryptographic framework of providing counter-measures against various classes of tampering attacks. This work was pioneered by the early works of [ISW03, GLM+03, IPSW06], and has since led to many subsequent models (see [KKS11, LL12] for an extensive discussion of these models).

No code can protect against a completely unrestricted adversary. For this reason, NMCs are only required to work for restricted families of "tampering functions" that an adversary may inflict. An NMC limits an adversary's control over the decoded message by introducing randomness in the encoding procedure, whereby the encoding function randomly samples a codeword from a distribution that depends on the message. More formally, for an alphabet Γ and a family of tampering functions $\mathcal{F}$ mapping Γ^n to itself, an NMC that protects against $\mathcal{F}$ consists of a randomized encoding function $\mathsf{Enc} : \{0,1\}^k \to \Gamma^n$, mapping messages to Γ^n-valued random variables, and a deterministic decoding function $\mathsf{Dec} : \Gamma^n \to \{0,1\}^k \cup \{\bot\}$, where $\bot$ denotes error detection. Roughly, the pair $(\mathsf{Enc}, \mathsf{Dec})$ satisfies the following property. For every $x \in \{0,1\}^n$ and for every $f \in \mathcal{F}$, the (random) codeword $X = \mathsf{Enc}(x)$ decodes correctly as $\mathsf{Dec}(X) = x$, but the corrupted version $Y = f(X)$ either decodes to x, or to a random variable $\mathsf{Dec}(Y)$ whose distribution is close to a distribution $\mathcal{D}_f$ depending on f but *not* on x.[1] The main goal is to design NMCs for large classes of tampering functions while maximizing the *rate* $k/(n \log |\Gamma|)$.

---

[1] We refer to [DPW10] for a more formal definition.

The class of tampering functions that has been studied most in the past literature arises in the so-called *split-state* model. In this model, the codeword index-set $[n]$ is partitioned into two roughly equally-sized sets $I_1, I_2 \subseteq [n]$ and the tampering functions consist of pairs $f = (f_1, f_2)$, where $f_i : \Gamma^{I_i} \to \Gamma^{I_i}$ is arbitrary. Codewords are then seen as being "split" into two states $X = (X_1, X_2)$, where $X_i \in \Gamma^{I_i}$, and tampered codewords have the form $Y = (f_1(X_1), f_2(X_2))$.

Aggarwal, Dodis, and Lovett [ADL14] gave the first and by far the simplest construction in the split-state model. For a prime number p and positive integer n, their encoding function sends $\{0,1\}^k$ into $\mathbb{F}_p^n \times \mathbb{F}_p^n$, giving split-state codewords of length $2n$ over the alphabet $\Gamma = \mathbb{F}_p$. Based on an improved construction of a so-called *affine-evasive set* due to [Agg15], their proof shows that the construction has the desired properties when $\log p = \Omega(k)$ and $n = \Omega(\log^6 p)$, which translates to a rate of roughly $\log^{-6} k$. However, the authors conjecture that their construction gives NMCs for constant n (independent of p), giving constant-rate codes. Although constant-rate split-state NMCs were later shown in [ADKO15], trying to search for the best-possible parameters for the [ADL14] construction is interesting for the following two reasons. First, the construction from [ADL14] is much simpler than the construction of [ADKO15], which was obtained by adding several layers of encodings to an already complex construction of Chattopadhyay and Zuckerman [CZ14]. Second, though the rate of [ADKO15] is a constant, this constant is very small and given the number of layers used in the construction, it is unlikely that it can be improved significantly. In contrast, there is no obvious reason why the construction of [ADL14, Agg15] cannot yield codes of rate $1/20$.

Towards determining the optimal parameters for the [ADL14] construction, we show that it still works when $n = C \log^4 p$ for a sufficiently large constant C, giving rate roughly $\log^{-4} k$. To this end, we improve a key element of the security proof of the construction, namely the following striking property of the inner-product function.

Theorem 1.1 (Aggarwal–Dodis–Lovett)**.** *There exist absolute constants $c, C \in (0, \infty)$ such that the following holds. Let p be a prime, $n \geq C \log^6 p$ be an integer, L, R be independent uniformly distributed random variables on $\mathbb{F}_p^n$ and $f, g : \mathbb{F}_p^n \to \mathbb{F}_p^n$ be functions. Then, there exist random variables u, a, b on $\mathbb{F}_p$ such that u is uniformly distributed, (a, b) is independent of u, and the distributions of $(\langle L, R \rangle, \langle f(L), g(R) \rangle)$ and $(u, au + b)$ have statistical distance at most $2^{-cn^{1/6}}$.*

The result roughly says that for any f, g, the random variable $\langle f(L), g(R) \rangle$ is correlated with some random variable of the form $au + b$. The restriction on n imposed in the theorem is directly responsible for the restriction on the codeword length in the [ADL14] construction. Improving this therefore implies higher-rate codes. The proof of Theorem 1.1 relies crucially on a breakthrough result of Sanders [San12] in additive combinatorics, concerning sumsets in general Abelian groups (see below). We improve Theorem 1.1 by simply exposing the dependence of Sanders's result on the magnitude of p when one restricts to the group $\mathbb{F}_p^n$. In particular, we obtain the following result (where all unspecified objects are as in Theorem 1.1).

Theorem 1.2. *There exist absolute constants $c, C \in (0, \infty)$ such that the following holds. Let $n \geq C \log^4 p$ be an integer and $f, g : \mathbb{F}_p^n \to \mathbb{F}_p^n$ be functions. Then, there exist random variables u, a, b on $\mathbb{F}_p$ such that u is uniformly distributed, (a, b) is independent of u, and the distributions of $(\langle L, R \rangle, \langle f(L), g(R) \rangle)$ and $(u, au + b)$ have statistical distance at most $2^{-cn^{1/4}}$.*

1.2. **The Quasi-polynomial Freĭman-Ruzsa Theorem.** For a finite Abelian group G and a subset $A \subseteq G$, define the *sum set* and *difference set* to be $A + A = \{a + b : a, b \in A\}$ and $A - A = \{a - b : a, b \in A\}$, respectively. The sizes of these sets are clearly bounded by $|A|^2$, but if A is a coset of a subgroup of G, then these sizes are exactly $|A|$. Conversely, if $|A \pm A| = |A|$, then A must be a coset of a subgroup. The identity $|A \pm A|/|A| = 1$ thus allows one to infer that A possesses a lot of structure. One of the most important conjectures in additive combinatorics,

the *Polynomial Freĭman-Ruzsa (PFR) Conjecture* (attributed to Marton in [Ruz99]), states that similar inferences can be made for sets in the group $\mathbb{F}_2^n$ that satisfy $|A + A|/|A| \ll |A|$.

Conjecture 1.3 (PFR Conjecture). Let $A \subseteq \mathbb{F}_2^n$ be such that $|A - A| \leq K|A|$. Then, there exists a set $B \subseteq A$ of size $|B| \geq |A|/C_1(K)$ that is contained in a coset of a subspace of size at most $C_2(K)|A|$, where $C_1(K)$ and $C_2(K)$ are polynomial in K.

This conjecture is sometimes stated differently in the literature; see [Gre] for the equivalence of five common formulations. The above formulation is the one which appears most frequently in Theoretical Computer Science, where it found several important applications, such as in linearity testing [Sam07], extractors [ZBS11, AHL15], error-correcting codes [BDL13] and communication complexity [BSLRZ14]. Major progress towards proving the PFR conjecture was made not long ago by Sanders [San12], whose result applies to general Abelian groups as opposed to just to $\mathbb{F}_2^n$. Recall that an Abelian group G has *torsion* r if $rg = 0$ for every $g \in G$.[2] The groups we care about here, namely $\mathbb{F}_p^n$, thus have torsion p. For groups of bounded torsion, Sanders's result implies the following [San12, Theorem 11.1].

Theorem 1.4 (Bogolyubov-Ruzsa Lemma for bounded torsion Abelian groups). *For every positive integer r there exists a $c(r) \in (0, \infty)$ such that the following holds. Let G be an Abelian group of torsion r, let $A, B \subseteq G$ be such that $|A + B| \leq K \min\{|A|, |B|\}$. Then, $(A - A) + (B - B)$ contains a subgroup $|V|$ of size at least $|V| \geq |A + B|/2^{c \log^4 2K}$.*

Standard arguments (see the proof of Lemma 1.6 in Section 2) show that Theorem 1.4 implies the *quasi-polynomial Freĭman-Ruzsa Theorem*: the statement of Conjecture 1.3 but with $C_1(K) = 2^{-C \log^4 K}$ and $C_2(K) = K^c$ for absolute constants $C, c \in (0, \infty)$. For NMCs, it was shown in [ADL14] that the following corollary of [San12] — an $\mathbb{F}_p$-analogue of the quasi-PFR Theorem [ADL14, Lemma 10] — implies Theorem 1.1.

Lemma 1.5. *There exist absolute constants $C, c \in (0, \infty)$ such that the following holds. Let p be a prime, n be a positive integer, $A \subseteq \mathbb{F}_p^n$ be such that $|A - A| \leq K|A|$. Then, there exists a set $B \subseteq A$ of size $|B| \geq |A|/p^{C \log^6 K}$ such that $\operatorname{Span}(B) \leq K^c|A|$.*

Our improvement over Theorem 1.1, Theorem 1.2, follows from the following variant of the above lemma, which we derive by exposing the dependence on the torsion p of $\mathbb{F}_p^n$ in Theorem 1.4.

Lemma 1.6. *Let $A \subseteq \mathbb{F}_p^n$ be such that $|A - A| \leq K|A|$. Then, there exists a set $B \subseteq A$ of size $|B| \geq |A|/p^{C \log^4(Kp)}$ such that $|\operatorname{Span}(B)| \leq pK^c|A|$, for absolute constants $C, c \in (0, \infty)$.*

This improves Lemma 1.5 if $K = \operatorname{poly}(p)$. Since this is the case for the application to NMCs, the proof of Theorem 1.1 given in [ADL14], but based on Lemma 1.6 instead of Lemma 1.5, gives Theorem 1.2.

1.3. **Linearity testing.** One further application of Lemma 1.6 is to linearity testing. The linearity test of Samoronidsky [Sam07] checks if a function $f : \mathbb{F}_p^n \mapsto \mathbb{F}_p^n$ is linear by picking $x, x' \in \mathbb{F}_p^n$ uniformly at random and accepting (declaring f linear) if and only if $f(x - x') = f(x) - f(x')$. Based on Lemma 1.5 it was shown in [Sam07, ADL14] that the test rejects if f is far from being linear. More precisely, it was shown that if $\Pr_{x,x' \in \mathbb{F}_p^n}[f(x - x') = f(x) - f(x')] \geq \varepsilon$, then there exists a linear function $g : \mathbb{F}_p^n \to \mathbb{F}_p^n$ such that $\Pr[f(x) = g(x)] \geq \varepsilon'$, where $\varepsilon' = p^{-O(\log^6(1/\varepsilon))}$ for an absolute constant $c \in (0, \infty)$. Using Lemma 1.6 in the proof instead results in a bound of $\varepsilon' = p^{-O(\log^4(p/\varepsilon))}$, which is an improvement if $\varepsilon = p^{-\Omega(1)}$.

---

[2]If the group operation is written multiplicatively, then the group is said to have exponent r.

2. Reduction to a Bogolyubov-Ruzsa Lemma for large sets

To avoid repeating the phrase *"for absolute constants"* many times, in the sequel C, C' will denote such constants whose value in $(0, \infty)$ may change from line to line. For the proof of Lemma 1.6 we closely follow Lovett's excellent exposition [Lov15] of the result for the group $\mathbb{F}_2^n$. The first step is a standard reduction to the following special case of Theorem 1.4.

Theorem 2.1. *Let $A \subseteq \mathbb{F}_p^n$ be such that $|A| \geq p^{n-1}/K^{24}$ and $|A - A| \leq K|A|$. Then, there exists a subspace $V \subseteq 2A - 2A$ of size $|V| \geq p^{-C \log^4(K \cdot p)}|A|$.*

The reduction uses the fact that Theorem 2.1 implies the following seemingly stronger version in which $|A|$ is not bounded.

Theorem 2.2. *Let $A \subseteq \mathbb{F}_p^n$ be such that $|A - A| \leq K|A|$. Then, there exists a subspace $V \subseteq 2A - 2A$ of size $|V| \geq |A|/p^{C \log^4(K \cdot p)}$.*

Lemma 1.6 follows by combining the above theorem with the following result of Plünnecke [Plü69].

Lemma 2.3 (Plünnecke). *Let G be an Abelian group and $A \subseteq G$ be such that $|A - A| \leq K|A|$. Then, for positive integers k, ℓ, we have $|kA - \ell A| \leq K^{k+\ell}|A|$.*

Proof of Lemma 1.6: Let $V \subseteq 2A - 2A$ be a subspace as in Theorem 2.2. Let $R \subseteq A$ be a maximal set such that no two elements from R belong to the same coset of V. By Lemma 2.3, $|R||V| = |R + V| = |A + V| \leq K^5|A|$. It follows that $|R| \leq K^5 p^{C \log^4(K \cdot p)}$. Let $B = A \cap (V + g)$ be such that $|B|$ is maximal. Then, $|B| \geq |A|/|R| \geq |V|/K^5 \geq |A|/K^5 p^{C \log^4(K \cdot p)}$ and moreover, $|\operatorname{Span}(B)| \leq |\operatorname{Span}(V + g)| \leq p|V| \leq pK^4|A|$. $\qquad\square$

Lemma 2.4. *Theorem 2.1 implies Theorem 2.2.*

The proof of Lemma 2.4 uses the following definition and proposition.

Definition 2.5. Let $A \subseteq \mathbb{F}_p^n$ be a subset and t, m be positive integers and $\phi : \mathbb{F}_p^n \to \mathbb{F}_p^m$ be a linear map. Then ϕ is a *Freiman homomorphism* of order t of the set A if for any $k, \ell \in \mathbb{N} \cup \{0\}$, $k + \ell = t$, it maps distinct elements $a, b \in kA - \ell A$ to distinct elements $\phi(a), \phi(b)$ in $\mathbb{F}_p^m$.

Proposition 2.6. *Let $A \subseteq \mathbb{F}_p^n$ be a subset and let t be a positive integer. Let m be the smallest integer such that there exists a Freiman homomorphism $\phi : \mathbb{F}_p^n \to \mathbb{F}_p^m$ of order t of the set A. Then, $\phi(\mathbb{F}_p \cdot (tA - tA)) = \mathbb{F}_p^m$.*

Proof: Towards a proof by contradiction, assume that there exists an $x \in \mathbb{F}_p^m$ such that for any $\alpha \in \mathbb{F}_p \setminus \{0\}$, $\alpha \cdot x \notin \phi(tA - tA)$. We will show that the existence of such a vector contradicts the assumption that m is minimal. To this end, let $\psi : \mathbb{F}_p^m \to \mathbb{F}_p^{m-1}$ be an arbitrary linear map whose kernel is the subspace S of $\mathbb{F}_p^m$ spanned by x. We show that $\psi \circ \phi : \mathbb{F}_p^n \to \mathbb{F}_p^{m-1}$ is also a Freiman homomorphism of order t of A, which implies our contradiction to the minimality of m.

Let a, b be arbitrary distinct vectors in $kA - \ell A$ for some $k, \ell \in \mathbb{N} \cup \{0\}$, $k + \ell = t$. Moreover, since (by assumption) ϕ is linear and in particular a Freiman homomorphism, and $S \cap \phi(tA - tA) = \{0\}$, we have $\phi(a) - \phi(b) = \phi(a - b) \notin S$. Hence, since the composition of two linear operators is again linear, $\psi \circ \phi(a) - \psi \circ \phi(b) = \psi\big(\phi(a - b)\big) \neq 0$, which is to say that $\psi \circ \phi$ maps distinct elements of $kA - \ell A$ to distinct elements in $\mathbb{F}_p^{m-1}$. In other words $\psi \circ \phi$ is a Freiman homomorphism of order t of A, giving the desired contradiction. $\qquad\square$

Proof of Lemma 2.4: Let $A \subseteq \mathbb{F}_p^n$ be a set such that $|A - A| \leq K|A|$ as in the statement of Theorem 2.2. We begin by observing that without loss of generality, we may assume that $0 \in A$. To see this, let $a \in A$ be an arbitrary element and consider the set $A' = A - a$, which certainly contains the origin. Also, $A' - A' = A - A$ and therefore $|A' - A'| = |A - A| \leq K|A| = K|A'|$, but

also $2A' - 2A' = 2A - 2A$. Hence, if there exists a subspace $V \subseteq 2A' - 2A'$ of the size promised by Theorem 2.2 then V is also a subspace of $2A - 2A$.

Notice that since the assumption $0 \in A$ implies that $\ell A \subseteq tA$ for every $\ell \in [t]$, it follows that a Freiman homomorphism of order t of A is also a Freiman homomorphism of order ℓ of A.

To deduce the conclusion of Theorem 2.2 from Theorem 2.1 we use a Freiman homomorphism to embed A into a smaller space such that it occupies a large fraction of it. To this end, let $\phi : \mathbb{F}_p^n \to \mathbb{F}_p^m$ be a Freiman homomorphism of order 12 of A such that m is minimal. Let $B = \phi(A) \subseteq \mathbb{F}_p^m$ be the embedding of A into $\mathbb{F}_p^m$. Notice that since ϕ is also a Freiman homohorphism of order 2 of A, we have $|B - B| = |A - A| \leq K|A| = K|B|$. Moreover, by Lemma 2.3 and Proposition 2.6,

$$p^m = |\mathbb{F}_p^m| = |\phi(\mathbb{F}_p \cdot (12A - 12A))| = |\mathbb{F}_p(12\phi(A) - 12\phi(A))| \leq K^{24} \cdot p \cdot |B|,$$

where the third equality uses linearity of ϕ. Thus $|B|$ satisfies the conditions of Theorem 2.1.

Theorem 2.1 says that there exists a subspace $V \subseteq 2B - 2B$ of size $|V| \geq p^{-C \log^4(K \cdot p)}|B|$. Since $|B| = |A|$, the subspace $|V|$ has the same size as the subspace promised to exist by Theorem 2.2. We now lift V into the larger space $\mathbb{F}_p^n$ such that it becomes contained in $2A - 2A$. To this end, recall that ϕ maps distinct elements of $2A - 2A$ into distinct elements of $2B - 2B$, since $2B - 2B = \phi(2A - 2A)$ and ϕ is a Freiman homomorphism of order 4 of A. Hence for each element $v \in V$ there is a unique element $v' \in 2A - 2A$ such that $\phi(v') = v$. Let $V' \subseteq 2A - 2A$ be the set of size $|V'| = |V|$ such that $\phi(V') = V$.

What is left is to show that V' is in fact a subspace of $\mathbb{F}_p^n$. Clearly it suffices to show that if $u', v' \in V'$ then $u' + v' \in V'$. Let $u = \phi(u')$ and $v = \phi(v')$. Since V is a linear subspace, $w = u + v \in V$. Let w' be the unique element in V' such that $\phi(w') = w$. Since $u', v,' w' \in 2A - 2A$ we have $u' + v' - w' \in 6A - 6A$. As ϕ is a Freiman homomorphims of order 12 of A and must map $0 \in 12A$ to $0 \in \mathbb{F}_p^m$, we have $\phi(u' + v' - w') = u + v - w = 0$. This implies $u' + v' - w' = 0$ and thus $u' + v' = w' \in V'$ as desired. $\qquad\square$

3. Proof of Theorem 2.1

The proof of Theorem 2.1 is split into two parts.

3.1. Part 1: A large gentle shifting set.

The first part uses the following standard notational conventions and definitions. For a set $A \subseteq \mathbb{F}_p^n$, let $\mathbf{1}_A : \mathbb{F}_p^n \mapsto \{0, 1\}$ be the indicator function of A, and let $\rho_A(x) = (p^n/|A|)\mathbf{1}_A(x)$ be its density function. For $a \in \mathbb{F}_p^n$, we shorthand $\rho_{\{a\}}(x)$ by $\rho_a(x)$. The convolution of two functions $f, g : \mathbb{F}_p^n \mapsto \mathbb{R}$ is defined by

$$f * g(x) = \mathbb{E}_{y \in \mathbb{F}_p^n}[f(y)g(x - y)]$$

and their inner product by

$$\langle f, g \rangle = \mathbb{E}_{x \in \mathbb{F}_p^n}[f(x)g(x)].$$

Note the identities $\rho_A * f(x) = \mathbb{E}_{a \in A}[f(x - a)]$ and $\rho_x * f(a) = f(x - a)$.

The main result of this section is as follows.

Lemma 3.1. *Let $A \subseteq \mathbb{F}_p^n$ be such that $|A| \geq p^n/L$. Then, for any $\tau > 0$, there exists a set $X \subseteq \mathbb{F}_p^n$ of size $|X| \geq p^n/(2L^{C\tau^2 \log^3 L})$ such that for every positive integer $t \leq \tau \log L$ and every $x \in tX$, we have $\mathbb{E}_{a,b \in A}[\mathbf{1}_{A-A}(a - b - x)] \geq 0.9$.*

The proof of the above lemma relies crucially on the following $\mathbb{F}_p^n$-version of a more general result of Croot and Sisask [CS10].

Lemma 3.2. *Let $A \subseteq \mathbb{F}_p^n$ be such that $|A| \geq p^n/L$ and let $f : \mathbb{F}_p^n \to [0, 1]$ be a function. Then, for any $q \geq 1$ and $\varepsilon \in (0, 1)$, there exists a set $X \subseteq \mathbb{F}_p^n$ of size $|X| \geq p^n/(2L^{q/\varepsilon^2})$ such that for every $x \in X$,*

$$\|\rho_x * \rho_A * f - \rho_A * f\|_q \leq C\varepsilon.$$

Proof: Set $\ell = q/\varepsilon^2$. We show that for each $x \in \mathbb{F}_p^n$, at least half of the ℓ-tuples $(a_1, \ldots, a_\ell) \in A^\ell$ satisfy

$$\tag{1} \left\|\rho_{(A+x)} * f - \frac{1}{\ell}\sum_{i=1}^{\ell} \rho_{a_i} * f\right\|_q \leq \frac{C\varepsilon}{2}.$$

To see why this suffices, for each $x \in \mathbb{F}_p^n$ let $S_x \subseteq (\mathbb{F}_p^n)^\ell$ be the set of ℓ-tuples in A^ℓ such that (1) holds. We have $|S_x| \geq |A|^\ell/2 \geq p^{n\ell}/(2 \cdot L^\ell)$. By the Pigeonhole Principle, there exists an ℓ-tuple $(a_1, \ldots, a_\ell) \in A^\ell$ that, for some set $X' \subseteq \mathbb{F}_p^n$ of size at least $p^n/(2 \cdot L^\ell)$, belongs to S_x for each $x \in X'$. By the triangle inequality, this implies that for every $x, x' \in X'$, we have

$$\|\rho_{A+x} * f - \rho_{A+x'} * f\|_q \leq \left\|\rho_{(A+x)} * f - \frac{1}{\ell}\sum_{i=1}^{\ell}\rho_{a_i} * f\right\|_q + \left\|\rho_{(A+x')} * f - \frac{1}{\ell}\sum_{i=1}^{\ell}\rho_{a_i} * f\right\|_q$$

$$\tag{2} \leq C\varepsilon.$$

Since the ℓ_q norm is invariant under permutations of the coordinates, the left-hand side of (2) equals

$$\|\rho_{A+x} * f - \rho_{A+x'} * f\|_q = \|\rho_{A+x+x'} * f - \rho_A * f\|_q.$$

For any $x \in X'$, the set $X = X' + x$ then has the desired properties since $|X| = |X'| \geq p^n/(2L^\ell)$.

We thus set out to prove that for every $x \in \mathbb{F}_p^n$, inequality (1) holds for at least half of A^ℓ. Let $a_1, \ldots, a_\ell$ be independent uniformly distributed A-valued random variables. For each $i \in [\ell]$ and $x, y \in \mathbb{F}_p^n$ define the random variable $Y_i^x(y) = \rho_{A+x} * f(y) - \rho_{a_i} * f(y)$. Notice that since the functions $\rho_{A+x} * f$ and $\rho_{a_i} * f$ are $[0,1]$-valued, we have $|Y_i^x(y)| \leq 2$. Hence, by definition of the ℓ_q norms, linearity of expectation and the definition of the $Y_i^x(y)$ random variables,

$$\mathbb{E}\left[\left\|\rho_{A+x} * f - \frac{1}{\ell}\sum_{i=1}^{\ell}\rho_{a_i} * f\right\|_q^q\right] = \mathbb{E}\left[\mathbb{E}_{y \in \mathbb{F}_p^n}\left[\left|\frac{1}{\ell}\sum_{i=1}^{\ell}\left(\rho_{A+x} * f(y) - \rho_{a_i} * f(y)\right)\right|^q\right]\right]$$

$$= \mathbb{E}_{y \in \mathbb{F}_p^n}\left[\mathbb{E}\left[\left|\frac{1}{\ell}\left(Y_1^x(y) + \cdots + Y_\ell^x(y)\right)\right|^q\right]\right]$$

$$\leq (C'q/\ell)^{q/2}$$

$$= (\sqrt{C'}\varepsilon)^q,$$

where the last two lines follow from the Marcinkiewicz-Zygmund inequality [MZ37] and our choice of ℓ. Hence, by Markov's inequality,

$$\tag{3} \Pr\left[\left\|\rho_{A+x} * f - \frac{1}{\ell}\sum_{i=1}^{\ell}\rho_{a_i} * f\right\|_q \leq \frac{\sqrt{C}\varepsilon}{2}\right] \geq \frac{1}{2},$$

showing that if we let $C = \sqrt{C'}$, at least half of the ℓ-tuples in A^ℓ satisfy (1) as required. $\square$

Proof of Lemma 3.1: Let $q \geq 1$ and $\varepsilon \in (0,1)$ be parameters to be set later and let $X \subseteq \mathbb{F}_p^n$ be a set as promised to exist by Lemma 3.2. Let t be a positive integer. We begin by observing that for every $x \in tX$,

$$\Pr_{a,b \in A}[a - b - x \in A - A] = \mathbb{E}_{a,b \in A}[\mathbf{1}_{A-A}(a - b - x)]$$

$$= \langle \rho_x * \rho_A * \mathbf{1}_{A-A}, \rho_A \rangle.$$

Moreover, by Hölder's inequality,

$$1 - \langle \rho_x * \rho_A * \mathbf{1}_{A-A}, \rho_A \rangle = \langle \rho_A * \mathbf{1}_{A-A} - \rho_x * \rho_A * \mathbf{1}_{A-A}, \rho_A \rangle$$

$$\tag{4} \leq \|\rho_A * \mathbf{1}_{A-A} - \rho_x * \rho_A * \mathbf{1}_{A-A}\|_q \|\rho_A\|_r,$$

6

where r is defined by $1/q + 1/r = 1$. To lower bound the above expectations we upper bound the two norms in (4). By our assumption on the size of A, we have $\|\rho_A\|_r = (p^n/|A|)^{(r-1)/r} \leq L^{1/q}$. Let $x = x_1 + \cdots + x_t$, where $x_i \in X$ for $1 \leq i \leq t$. To bound the remaining norm we telescope the difference

$$\rho_A * \mathbf{1}_{A-A} - \rho_{x_1 + \cdots + x_t} * \rho_A * \mathbf{1}_{A-A} = \sum_{i=1}^{t} (\rho_{x_1 + \cdots + x_{i-1}} * \rho_A * \mathbf{1}_{A-A} - \rho_{x_1 + \cdots + x_i} * \rho_A * \mathbf{1}_{A-A}).$$

Then, by the triangle inequality, invariance of the ℓ_q norm under permutations of the coordinates and Lemma 3.2 we get

$$\|\rho_A * \mathbf{1}_{A-A} - \rho_{x_1 + \cdots + x_t} * \rho_A * \mathbf{1}_{A-A}\|_q \leq \sum_{i=1}^{t} \|\rho_{x_1 + \cdots + x_{i-1}} * \rho_A * \mathbf{1}_{A-A} - \rho_{x_1 + \cdots + x_i} * \rho_A * \mathbf{1}_{A-A}\|_q$$

$$= \sum_{i=1}^{t} \|\rho_A * \mathbf{1}_{A-A} - \rho_{x_i} * \rho_A * \mathbf{1}_{A-A}\|_q \leq tC\varepsilon.$$

Hence, for every $x \in tX$, we have

$$\mathbb{E}_{a,b}[1_{2A}(x + a + b)] \geq 1 - \|\rho_A * \mathbf{1}_{A-A} - \rho_x * \rho_A * \mathbf{1}_{A-A}\|_q \|\rho_A\|_r \geq 1 - tC\varepsilon \cdot L^{1/q}.$$

Set $\varepsilon = 1/(20Ct)$, $q = \log L$. Then, for $t \leq \tau \log L$ the above bound becomes 0.9 and the size of X is at least $|X| \geq p^n/(2L^{q/\varepsilon^2}) \geq p^n/(2L^{400C^2\tau^2 \log^3 L})$. $\qquad\square$

3.2. **Part 2: A subspace from the Fourier support.** Denote $\omega = e^{2\pi i/p}$. For $X \subset \mathbb{F}_p^n$, $u \in \mathbb{F}_p^n$, and $\gamma \in [0,1]$, define

$$\widehat{X}(u) = \mathbb{E}_{x \in X}[\omega^{\langle u, x \rangle}]$$

and define

$$\mathrm{Spec}_\gamma(X) = \{u \in \mathbb{F}_p^n : |\widehat{X}(u)| \geq \gamma\}.$$

For $Y \subseteq \mathbb{F}_p^n$, let

$$Y^\perp = \{v \in \mathbb{F}_p^n : \langle u, v \rangle = 0, \quad \forall u \in Y\}.$$

Theorem 3.3 (Chang [Cha02]). *Let $X \subseteq \mathbb{F}_p^n$. Then*

$$\dim(\mathrm{Spec}_\gamma(X)) \leq 8\gamma^{-2} \log(p^n/|X|).$$

Lemma 3.4. *Let t be a positive integer, $A, X \subseteq \mathbb{F}_p^n$. Let $a, b \in A$, $x_1, \ldots, x_t \in X$ be independent uniformly distributed random variables and let $x = x_1 + \cdots + c_t$. Assume that*

$$(5) \qquad \mathbb{E}[\mathbf{1}_{A-A}(a - b - x)] \geq 1 - \varepsilon.$$

Set $V = \mathrm{Span}(\mathrm{Spec}_{1/2}(X))^\perp$. Then, for an independent and uniformly distributed $v \in V$, we have

$$(6) \qquad \mathbb{E}[\mathbf{1}_{A-A}(a - b - x + v)] \geq 1 - \varepsilon - \frac{1}{2^t}\frac{p^n}{|A|}.$$

Proof: We will show that

$$(7) \qquad \left| \mathbb{E}[\mathbf{1}_{A-A}(a - b - x) - \mathbb{E}[\mathbf{1}_{A-A}(a - b - x + v)] \right| \leq \frac{p^n}{2^t|A|}.$$

To this end, we begin by observing that by the Fourier Inversion Formula, the first and second expectation in (7) can be written as

$$\sum_{u \in \mathbb{F}_p^n} \widehat{A}(u)\widehat{A}(-u)\widehat{X}(-u)^t \widehat{\mathbf{1}}_{A-A}(u) \qquad \text{and} \qquad \sum_{u \in \mathbb{F}_p^n} \widehat{A}(u)\widehat{A}(-u)\widehat{X}(-u)^t \widehat{V}(u)\widehat{\mathbf{1}}_{A-A}(u),$$

respectively.

Since V is a subspace, $\widehat{V}(u) = 1$ if $u \in V^\perp$ and $\widehat{V}(u) = 0$ otherwise. The left-hand side of (7) therefore becomes

$$\Big| \sum_{u \notin V^\perp} \widehat{A}(u)\widehat{A}(-u)\widehat{X}(-u)^t \widehat{\mathbf{1}}_{A-A}(u) \Big|.$$

By definition of V we have $V^\perp = \mathrm{Span}(\mathrm{Spec}_{1/2}(X))$. It follows that if $u \notin V^\perp$ then $|\widehat{X}(-u)| < 1/2$. Also it is easy to see that $|\widehat{\mathbf{1}}_{A-A}(u)| \leq 1$. Hence,

$$
\begin{aligned}
\Big| \sum_{u \notin V^\perp} \widehat{A}(u)\widehat{A}(-u)\widehat{X}(-u)^t \widehat{\mathbf{1}}_{A-A}(u) \Big| &\leq \frac{1}{2^t} \sum_{u \notin V^\perp} |\widehat{A}(u)\widehat{A}(-u)| \\
&\leq \frac{1}{2^t} \sum_{u \in \mathbb{F}_p^n} |\widehat{A}(u)| \cdot |\widehat{A}(-u)| \\
&\leq \frac{1}{2^t} \cdot \frac{p^n}{|A|} \,,
\end{aligned}
$$

which gives the result. $\qquad\square$

Proof of Theorem 2.1: Recall that A has size at least $|A| \geq 2^n/L$ for $L = K^{24} \cdot p$. Lemma 3.1 says that there exists a set $X \subseteq \mathbb{F}_p^n$ of size $|X| \geq p^n/(2L^{100C \log^3 L})$ such that for any $t \leq 10 \log L$ and any $x \in tX$, we have $\mathbb{E}_{a,b \in A}[\mathbf{1}_{A-A}(a - b - x)] \geq 0.9$. Theorem 3.3 and Lemma 3.4 combined give us a subspace V of size at least $|V| \geq p^n/(2^{32}p^{C' \log^4 L})$ where $C' = 3200C$, such that the left-hand side of (5) can be written as and bounded by

$$(8) \qquad\qquad \mathbb{E}\Big[\mathbb{E}_{v \in V}[\mathbf{1}_{A-A}(a - b - x + v)] \Big] \geq 0.8,$$

where the outer expectation is taken over uniformly distributed $a, b \in A$ and $x_1, \ldots, x_t \in X$.

We show that $V \subseteq 2A - 2A$, implying the result. By the averaging principle, (8) shows that there exist $a, b \in A$ and $x \in tX$ such that 80% of the $v \in V$ satisfy $a - b - x + v \in A - A$. This implies that $v \in A - A - a + b + x$. We use the pigeon hole principle to show that this implies that $V \subseteq 2A - 2A$. To this end, fix a $v \in V \backslash \{0\}$ and notice that there exists a complete matching of $|V|/2$ pairs $(v_1, v_2) \in V \times V$ such that $v = v_1 + v_2$ (take any $v_1 \in V$ and set $v_2 = v - v_1$). This implies that their exists $v_1, v_2 \in A - A - a + b + x$ such that $v_1 + v_2 = v$, which in turn implies that $v \in 2A - 2A$. $\qquad\square$

References

[ADKO15] D. Aggarwal, Y. Dodis, T. Kazana, and M. Obremski. Non-malleable reductions and applications. In *The 47th ACM Symposium on Theory of Computing (STOC)*. 2015.

[ADL14] D. Aggarwal, Y. Dodis, and S. Lovett. Non-malleable codes from additive combinatorics. In *STOC*. ACM, 2014.

[Agg15] D. Aggarwal. Affine-evasive sets modulo a prime. *Information Processing Letters*, 115(2):382–385, 2015.

[AHL15] D. Aggarwal, K. Hosseini, and S. Lovett. Affine-malleable extractors, spectrum doubling, and privacy amplification. *IACR Cryptology ePrint Archive*, 2015:1094, 2015.

[BDL13] A. Bhowmick, Z. Dvir, and S. Lovett. New bounds for matching vector families. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 823–832. ACM, 2013.

[BSLRZ14] E. Ben-Sasson, S. Lovett, and N. Ron-Zewi. An additive combinatorics approach relating rank to communication complexity. *Journal of the ACM (JACM)*, 61(4):22, 2014.

[CDTV16] S. Coretti, Y. Dodis, B. Tackmann, and D. Venturi. Non-malleable encryption: Simpler, shorter, stronger. In *Theory of Cryptography - 13th International Conference, TCC 2016-A, Tel Aviv, Israel, January 10-13, 2016, Proceedings, Part I*, pages 306–335. 2016.

[CGM+15] N. Chandran, V. Goyal, P. Mukherjee, O. Pandey, and J. Upadhyay. Block-wise non-malleable codes. *IACR Cryptology ePrint Archive*, 2015:129, 2015.

[Cha02]   M.-C. Chang. A polynomial bound in Freĭman's theorem. *Duke Mathematical Journal*, 113(3):399–420, 2002.

[CMTV15]   S. Coretti, U. Maurer, B. Tackmann, and D. Venturi. From single-bit to multi-bit public-key encryption via non-malleable codes. In *Theory of Cryptography - 12th Theory of Cryptography Conference, TCC 2015, Warsaw, Poland, March 23-25, 2015, Proceedings, Part I*, pages 532–560. 2015.

[CS10]   E. Croot and O. Sisask. A probabilistic technique for finding almost-periods of convolutions. *Geometric and functional analysis*, 20(6):1367–1396, 2010.

[CZ14]   E. Chattopadhyay and D. Zuckerman. Non-malleable codes in the constant split-state model. *FOCS*, 2014.

[DPW10]   S. Dziembowski, K. Pietrzak, and D. Wichs. Non-malleable codes. In *ICS*, pages 434–452. Tsinghua University Press, 2010.

[GLM⁺03]   R. Gennaro, A. Lysyanskaya, T. Malkin, S. Micali, and T. Rabin. Algorithmic Tamper-Proof (ATP) security: Theoretical foundations for security against hardware tampering. In M. Naor, editor, *First Theory of Cryptography Conference — TCC 2004*, volume 2951 of *LNCS*, pages 258–277. Springer-Verlag, February 19–21 2003.

[GPR15]   V. Goyal, O. Pandey, and S. Richelson. Textbook non-malleable commitments. *IACR Cryptology ePrint Archive*, 2015:1178, 2015.

[Gre]   B. J. Green. Notes on the Polynomial Freĭman-Ruzsa Conjecture. Expository note. Available at http://people.maths.ox.ac.uk/greenbj/papers/PFR.pdf.

[IPSW06]   Y. Ishai, M. Prabhakaran, A. Sahai, and D. Wagner. Private circuits II: Keeping secrets in tamperable circuits. In S. Vaudenay, editor, *Advances in Cryptology—EUROCRYPT 2006*, volume 4004 of *LNCS*, pages 308–327. Springer-Verlag, 2006.

[ISW03]   Y. Ishai, A. Sahai, and D. Wagner. Private circuits: Securing hardware against probing attacks. In D. Boneh, editor, *Advances in Cryptology—CRYPTO 2003*, volume 2729 of *LNCS*. Springer-Verlag, 2003.

[KKS11]   Y. T. Kalai, B. Kanukurthi, and A. Sahai. Cryptography with tamperable and leaky memory. In *Advances in Cryptology–CRYPTO 2011*, pages 373–390. Springer, 2011.

[LL12]   F.-H. Liu and A. Lysyanskaya. Tamper and leakage resilience in the split-state model. In *Advances in Cryptology–CRYPTO 2012*, pages 517–532. Springer, 2012.

[Lov15]   S. Lovett. *An Exposition of Sanders' Quasi-Polynomial Freiman-Ruzsa Theorem*. Number 6 in Graduate Surveys. Theory of Computing Library, 2015.

[MZ37]   J. Marcinkiewicz and A. Zygmund. Quelques théoremes sur les fonctions indépendantes. *Fund. Math*, 29:60–90, 1937.

[Plü69]   H. Plünnecke. Eigenschaften und abschätzungen von wirkungsfunktionen, bmwf, 1969.

[Ruz99]   I. Ruzsa. An analog of Freĭman's theorem in groups. *Astérisque*, 258(199):323–326, 1999.

[Sam07]   A. Samorodnitsky. Low-degree tests at large distances. In *ACM symposium on Theory of computing*, pages 506–515. ACM, 2007.

[San12]   T. Sanders. On the Bogolyubov–Ruzsa lemma. *Analysis & PDE*, 5(3):627–655, 2012.

[ZBS11]   N. Zewi and E. Ben-Sasson. From affine to two-source extractors via approximate duality. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 177–186. ACM, 2011.

SCHOOL OF COMPUTER AND COMMUNICATION SCIENCES, EPFL, INF 230, STATION 14, CH-1015, LAUSANNE, SWITZERLAND

E-mail address: `divesh.aggarwal@epfl.ch`

CWI, SCIENCE PARK 123, 1098 XG AMSTERDAM, THE NETHERLANDS

E-mail address: `j.briet@cwi.nl`